Starter Question

13 Anna investigated the CO emissions for Fords first registered in London in 2016 using data from the Large Data Set.

She calculated the mean of the CO emissions for **all** of these cars driven by male and by female drivers.

	Male	Female
$\bar{x}$	0.401	0.385
Number of drivers	33	37

13 (a) Find the mean CO emissions for the male and female driver data combined.

[2 marks]

$$\frac{(0.401 \times 33) + (0.385 \times 37)}{33 + 37} = 0.393\,(3\,sf)$$

Starter Question

13 (b) Virat also investigated CO emissions for Fords first registered in London in 2016. He correctly calculated the mean CO emissions for all of these cars to be 0.38

Using your knowledge of the Large Data Set, explain why there is a difference between the mean you calculated in (a) and Virat's result.

rge Data Set\Lesson 1\AQA-AS-A-MATHS-LDS-2019-2020.xlsx

[1 mark]

Virat has included company car

13 (c) Anna claims that the mean **carbon dioxide** emissions for Fords first registered in London in 2016 is 0.485 g/km.

Using your knowledge of the Large Data Set, explain why this value must be incorrect.

[1 mark]

CO2 emissions are in the 10s/10

L Data presentation and interpretation

L3

Interpret measures of central tendency and variation, extending to standard deviation.

Be able to calculate standard deviation, including from summary statistics.

Students should be able to:

- find the mean, median, mode, range, quartiles and interquartile range from data given in graphical or tabular form
- interpret values of the mean, median and mode and recognise these as measures of central tendency
- calculate standard deviation (or variance) using a calculator or from summary statistics of the form $\sum x$, $\sum x^2$ or $\sum (x-\bar{x})^2$
- recognise the standard deviation, variance, range and interquartile range as measures of variation.

L Data presentation and interpretation

Notes

- Students are expected to use a calculator's statistical functions to find **all** statistics for a set of data presented as a list or in a frequency table, including estimating for grouped data.
- Linear interpolation of median and quartiles for grouped data is not required.
- Whilst an informal understanding that there are two different values for standard deviation given on a calculator is useful, this specification does not formally address the estimation of population parameters by sample statistics. Thus, the formula we will use for standard deviation is $\sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$ (which is given in the formulae book), but students will not be penalised for using the unbiased estimator of the population standard deviation.
- Use of any particular symbol (s or σ) for standard deviation in statistics will be avoided in exam questions, because of the potential for confusion. Students will need to recognise the correct value of standard deviation on a calculator.
- For small data sets, the positions of the median and quartiles are usually given by $\frac{n+1}{4}, \frac{n+1}{2}, \frac{3(n+1)}{4}$ and it will often be convenient to ensure $n+1$ is a multiple of 4
- However, the quartiles are more relevant to large sets of data and here it is usually more convenient to replace $n+1$ by n
- Whenever possible, students should use a calculator to determine quartiles.
- Understand that values of statistics are estimates of the corresponding population parameters.

9.2 Central Tendency & Spread

Measures of Dispersion

Range □ largest value of data – smallest value of data

Interquartile Range (IQR) □ this is the spread of the middle 50% of data. To calculate IQR, find the first and third quartiles, Q_1 and Q_3, then IQR $= Q_3 - Q_1$.

Q_1 and Q_3 can be found by arranging the data into ascending order. The position of Q_1 is . The position of Q_3 is . The values of Q_1 and Q_3 can be read off the list of data in a similar way to the

9.2 Central Tendency & Spread

Measures of Dispersion

Variance () □ this is a measure of how spread out the data values are from the mean. The variance of n observations with a mean of can be calculated using:

9.2 Central Tendency & Spread

<u>Measures of Dispersion</u>

Standard Deviation (or s) □ this is just the square root of the variance. Therefore the standard deviation is:

9.2 Central Tendency & Spread

This is the formula as it appears on the formula sheet in the

Standard deviation

$$\sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

The variance is just the bit inside the square root sign.

9.2 Central Tendency & Spread

Consider this list of randomly generated numbers:

55, 65, 67, 46, 38, 53, 08, 63, 71, 91

For each item of data, subtract the mean and write down the differences, e.g.

55 – 55.7 = -0.7
65 – 55.7 = 9.3
67 – 55.7 = 11.3
46 – 55.7 = -9.7
38 – 55.7 = -17.7
53 – 55.7 = -2.7
08 – 55.7 = -47.7
63 – 55.7 = 7.3
71 – 55.7 = 15.3
91 – 55.7 = 35.3

9.2 Central Tendency & Spread

Find the total of all the differences. What do you get?

-0.7 + 9.3 + 11.3 + -9.7 + -17.7 + -2.7 + -47.7 + 7.3 + 15.3 + 35.3 = 0

If is each data item, what we've just done is .

These differences tell you how far away each data item is away from the mean.

If we don't square those differences, we'll always get zero!

9.2 Central Tendency & Spread

Squaring the differences from the mean will remove all the negatives.

If we then divide by and square root, we will get the ‘average’ deviation of each data point from the mean, or…
the STANDARD DEVIATION!

9.2 Central Tendency & Spread

Example 1

Consider the following data:
2, 5, 3, 11, 6, 8, 3, 8, 1, 6, 2, 23, 9, 11, 18, 19, 22, 7.

Find:

a) The range

23 – 1 = **22**

b) The IQR

n = 18   1, 2, 2, 3, 3, 5, 6, 6, 7, 8, 8, 9, 11, 11, 18, 19, 22, 23

Position of Q_1 = = 4.75 5^{th} value   **Q_1 = 3**

Position of Q_3 = = 14.25 14^{th} value **Q_3 = 11**

From the calc:

9.2 Central Tendency & Spread

Example 1

1, 2, 2, 3, 3, 5, 6, 6, 7, 8, 8, 9, 11, 11, 18, 19, 22, 23

Find:

c) The variance

From the calc:

Standard deviation

$$\sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

Either is valid.

So variance

9.2 Central Tendency & Spread

Example 1

1, 2, 2, 3, 3, 5, 6, 6, 7, 8, 8, 9, 11, 11, 18, 19, 22, 23

Find:

d) The standard deviation

From the calc:

$$\sqrt{\frac{2322}{18}-\left(\frac{82}{9}\right)^2}=6.78\,(3\,sf)$$

Either is valid.

9.2 Central Tendency & Spread

Example 2

Find the variance, standard deviation, range and interquartile range of the data in the

x	frequency
2	2
3	5
4	5
5	4
6	1
7	1

Variance:

Standard Deviation:

Range: Max – Min

Interquartile Range:

9.2 Central Tendency & Spread

<u>Example 3</u>

The heights of sunflowers in a garden were measured and recorded in the table below:

Height, h cm	f
150 ≤ h < 170	5
170 ≤ h < 190	10
190 ≤ h < 210	12
210 ≤ h < 230	3

Remember to enter the mid-points as this is grouped data.

Variance:
Standard Deviation:
Largest range:

... riance and the standard ... heights. Find the largest possible range.

9.2 Central Tendency & Spread

Example 4

In each of the following situations calculate the standard deviation from the given summary statistics:

a) A sample of 50 pensioners were asked how many alcoholic drinks they consumed on a given evening. The results a

Standard deviation

$$\sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

$$sd = \sqrt{\frac{\sum\left(x-\bar{x}\right)^2}{n}} = \sqrt{\frac{68.4}{50}} = 1.368$$

9.2 Central Tendency & Spread

Example 4

b) A census of 1000 villagers is taken to gather information regarding the spread of their ages. The ages, x, are summarised as follows:

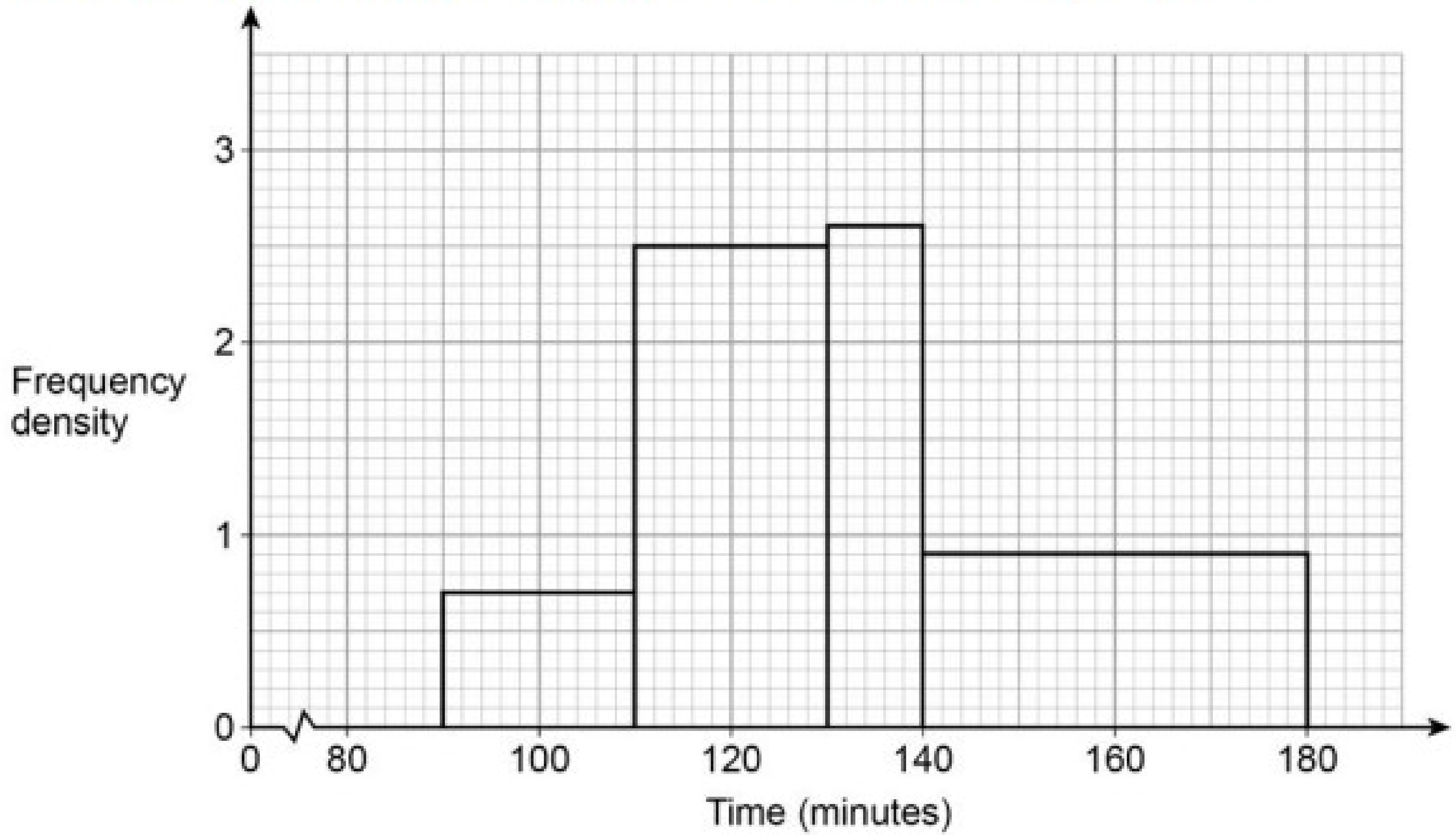
Standard deviation

$$\sqrt{\frac{\Sigma(x-\bar{x})^2}{n}} = \sqrt{\frac{\Sigma x^2}{n} - \bar{x}^2}$$

Note: if we needed the variance we just wouldn't square root

Find an estimate of the mean, standard deviation and variance from this histogram:

Amir is a sales manager who drives 100km to get to work. He times the journeys he makes in a six-month period and represents the data in the histogram below.



9.2 Central Tendency & Spread

Make a table and enter it on the calc…

We need the mid-point and the frequency before we can do any calculations.

From GCSE, frequency = frequency density x class width!

Minutes	Mid-point	Class Width	Frequency Density	Frequency
90 – 110	**100**	20	0.7	**14**
110 – 130	**120**	20	2.5	**50**
130 – 140	**135**	10	2.6	**26**
140 – 180	**160**	40	0.9	**36**

$\square = 132.3$ $\square^2 = 396.7$ $\square = 19.9$